

Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*

Mark A. Beilstein^{a,1}, Nathalie S. Nagalingum^{a,2}, Mark D. Clements^a, Steven R. Manchester^b, and Sarah Mathews^{a,1}

^aThe Arnold Arboretum of Harvard University, Jamaica Plain, MA 02138; and ^bFlorida Museum of Natural History, University of Florida, Gainesville, FL 32611

Edited* by Michael J. Donoghue, Yale University, New Haven, CT, and approved September 7, 2010 (received for review August 26, 2009)

Dated molecular phylogenies are the basis for understanding species diversity and for linking changes in rates of diversification with historical events such as restructuring in developmental pathways, genome doubling, or dispersal onto a new continent. Valid fossil calibration points are essential to the accurate estimation of divergence dates, but for many groups of flowering plants fossil evidence is unavailable or limited. *Arabidopsis thaliana*, the primary genetic model in plant biology and the first plant to have its entire genome sequenced, belongs to one such group, the plant family Brassicaceae. Thus, the timing of *A. thaliana* evolution and the history of its genome have been controversial. We bring previously overlooked fossil evidence to bear on these questions and find the split between *A. thaliana* and *Arabidopsis lyrata* occurred about 13 Mya, and that the split between *Arabidopsis* and the *Brassica* complex (broccoli, cabbage, canola) occurred about 43 Mya. These estimates, which are two- to threefold older than previous estimates, indicate that gene, genomic, and developmental evolution occurred much more slowly than previously hypothesized and that *Arabidopsis* evolved during a period of warming rather than of cooling. We detected a 2- to 10-fold shift in species diversification rates on the branch uniting Brassicaceae with its sister families. The timing of this shift suggests a possible impact of the Cretaceous–Paleogene mass extinction on their radiation and that Brassicales codiversified with pierid butterflies that specialize on mustard-oil-producing plants.

Bayesian dating | Brassica | Brassicaceae | chronogram | fossil calibration

The most important genetic model in plant biology is *Arabidopsis thaliana*. It is the first plant to have its entire genome sequenced, and it serves as a key comparison point with other eukaryotic genomes. *A. thaliana* is diploid and has a small genome distributed on just five chromosomes, considerations in its choice as a model (1). The age of the *Arabidopsis* crown group (CG), previously estimated at 5.8–3 Mya (2, 3), and of splits within Brassicaceae have been used to understand the pace of evolution in genes affecting self-incompatibility (4, 5), the rate of change in signal transduction and gene expression (6, 7), the persistence of shared chromosomal rearrangements in *A. thaliana* and *Brassica oleracea* (8), the tempo of evolution of miRNA sequences (9), the evolution of pierid butterflies specializing in plants that produce mustard oils (10), and the ages of whole-genome duplication (WGD) events giving rise to gene pairs in *Arabidopsis* (11). As genomes of additional Brassicaceae (e.g., *Capsella rubella*) and other Brassicales (e.g., *Carica papaya*) (12) are sequenced, the importance of robust estimates of divergence dates relating these genomes to one another and to the geological record increases substantially.

The accuracy of divergence times inferred from sequence data depends on valid, verifiable fossils to calibrate phylogenetic trees. Previous dates for the origin of *Arabidopsis* relied on the report of fossil pollen assigned to the genus *Rorippa* (Brassicaceae) (2, 3). We discovered that this report is not linked to a physical specimen, published image, or description; thus, its validity for calibration of a Brassicaceae phylogeny cannot be evaluated. Reexamination of fossils from the order Brassicales (Table S1) revealed six fossil taxa that are sufficiently documented to serve

as age constraints; however, only one has been used in previous estimations of divergence times (13). Among the taxa that had been overlooked is *Thlaspi primaevum* (Brassicaceae) (14), an Oligocene fossil with angustiseptate winged fruits (Fig. 1A) from the Ruby Basin Flora of southwestern Montana, dated at 30.8–29.2 Mya (15). It may have been overlooked because many of the generic determinations from the Ruby Basin Flora remain to be verified, or there may have been concerns about convergence of winged fruits from unrelated angiosperm families. Although the placement of *T. primaevum* in Brassicaceae now has been confirmed (16), this fruit form evolved independently multiple times within the family (17). However, among extant Brassicaceae, angustiseptate fruit combined with concentrically striate seeds (Fig. 1B, C, and E) is unique to species of *Thlaspi*. We examined the seed chamber of *T. primaevum* and found seed striations in the same pattern as those of extant *Thlaspi* seeds (Fig. 1D). *Alliaria petiolata* is the only other Brassicaceae with striated seeds (18) (Fig. S1), but it has longitudinally oriented striations, and the fruit is latiseptate; latiseptate fruits are ancestral for the clade defined by the coalescence of *A. petiolata* and *Thlaspi arvense* (Fig. S1). Thus, *T. primaevum* is a valid age constraint within CG Brassicaceae, and it was placed at the coalescence of *T. arvense* and *A. petiolata* as a minimum age constraint for this split (Fig. 2 and Figs. S2–S4).

The other five potentially useful Brassicales fossils are placed outside the Brassicaceae crown group. *Cappariodoxylon holleisii* is a Miocene wood fossil with affinities to extant *Capparis*, dated at 17–16.3 Mya (Table S1) (19). We explored its use as a constraint within Core Brassicales (Brassicaceae + Cleomaceae + Capparaceae) (*SI Materials and Methods*). *Dressiantha bicarpellata* is a Turonian fossil dated at 93.6–89.3 Mya (20), the oldest known putative brassicalean fossil; it provided the single age constraint used by Couvreur et al. (13). However, membership of *D. bicarpellata* within Brassicales is contentious (21), and thus its validity as an age constraint for estimating divergence times in the order remains unclear. To explore these issues we (i) assessed the impact of using it to constrain four nodes along the backbone of the Brassicales tree, based on results from analyses of combined nucleotide and morphological data (*SI Materials and Methods*) and (ii) estimated ages in Brassicales without *D. bicarpellata*. Finally, three species of *Akania* were used to constrain the ages of nodes within the *Bretschneidera/Akania*

Author contributions: M.A.B. and S.M. designed research; M.A.B., N.S.N., and S.R.M. performed research; M.D.C. contributed new reagents/analytic tools; M.A.B., N.S.N., M.D.C., and S.R.M. analyzed data; and M.A.B. and S.M. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: All accession numbers for sequences downloaded from GenBank and sequences deposited in GenBank as part of this study appear in Table S2.

¹To whom correspondence may be addressed. E-mail: mbeilstein@tamu.edu or smathews@oeb.harvard.edu.

²Present address: University of California Museum of Paleontology, University of California, Berkeley, CA 94720.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0909766107/-DCSupplemental.

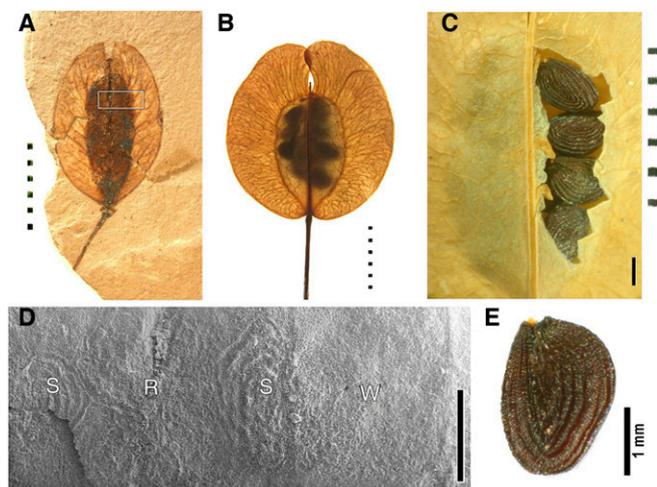


Fig. 1. Fossil and extant angustiseptate winged *Thlaspi* fruits and striated seeds. (A) *T. primaevum* fossil from the Ruby Basin Flora, western Montana. Extant *T. arvense* fruit backlit to show placement of seeds in the two locules (wings) (B) and with a portion of the valve removed to show striated seeds (C). Locules are separated by the replum. (D) Scanning electron micrograph of the fossil seed chamber (indicated by rectangle in A) showing impressions of striated seeds. (E) *T. arvense* striated seed. S, seed; R, replum; W, wing. (Dotted scale bars, 5 mm; solid scale bars, 1 mm.)

Tropaeolum clade (Fig. 2). To take full advantage of these age constraints deeper in Brassicales, we expanded our sample of the plastid locus NADH dehydrogenase subunit F (*ndhF*) and the nuclear locus phytochrome A (*PHYA*), which have provided a robust phylogenetic framework for the Brassicaceae (17, 22), to include members of the Core Brassicales, giving us a matrix of 179 species (*SI Material and Methods* and *Table S2*). We estimated divergence times for trees inferred from each locus and from the combined data.

Most estimates of divergence times in Brassicaceae have assumed that rates of nucleotide evolution are equal across the tree and have been based on either the mutation rate at synonymous sites (2, 3) or the estimated times of genome duplications (23–26). Two recent analyses have allowed rates to be uncorrelated across the tree, but one of these analyses relied on *D. bicarpellata* as a single age constraint (13), and the other relied on a single secondary age constraint (27). Neither of these analyses took into account uncertainties in the topology of the tree (13, 27). Our analyses used multiple age constraints, allowed rates of nucleotide evolution to be uncorrelated across the tree, and accounted for uncertainty in the phylogenetic hypothesis and in the placement of the fossil age constraints.

Results and Discussion

Using Bayesian approaches (28), we estimated the origin of CG *Arabidopsis* at ≈ 13.0 Mya [95% highest probability density (HPD): 17.9–8.0] (Fig. 2 and *Table 1*), considerably older than the frequently cited estimate of 5.8–3 Mya (*Table 1*) (2, 3). When a penalized likelihood approach (29, 30) was used, the estimate for this node was even older (*Table S3*). In the subsequent discussion, we focus on estimates from the Bayesian analyses (*Table 1* and *Table S3*). The placement or exclusion of *D. bicarpellata* had little impact on the age of this node and other nodes in the tree (*Table S3*). Conversely, the *T. primaevum* constraint did lead to slightly older ages for most nodes in the tree (*Table S4*). Nonetheless, even without this constraint within CG Brassicaceae, age estimates were substantially older than previous estimates (*Table 1*), possibly resulting from our use of multiple constraints, dense taxon sampling (but see ref. 13), from our

allowing for phylogenetic uncertainty, alternative placement of fossil calibrations, from the relative completeness of our molecular datasets (cf. ref. 13), or from a combination of these factors. The previous estimate for CG *Arabidopsis* is primarily Pliocene (5.3–1.8 Mya), whereas the revised age falls within the Miocene (23.03–5.3 Mya) and spans a particularly warm period in recent earth history that included the Middle Miocene climatic optimum (31). Thus, warming may have played a role in the divergence of *A. thaliana* from other *Arabidopsis*. Moreover, the much older optimal age estimate of 13.0 Mya suggests that the pace of chromosomal rearrangements (1), divergent gene regulation (6), and the breakdown of self-incompatibility (4) may have proceeded more slowly than has been appreciated.

The first split within CG Brassicaceae occurred in the early Eocene, ≈ 54.3 Mya (95% HPD: 64.2–45.2) (Fig. 2 and *Figs. S2–S4*). CG Lineages I, II, and III, which contain the majority of Brassicaceae (17, 22), radiated in the mid to late Eocene, from 43.4 to 33.3 Mya (*Table 1*). Because *A. thaliana* and *B. oleracea* (broccoli and related plants) occur in Lineages I and II, respectively, the relatively deep coalescence of these scientifically and economically important species dates to ≈ 43.2 Mya (95% HPD: 50.7–36.6; *Table 1*), indicating that conservation of the large chromosomal blocks shared by these two species (8) has persisted for longer than previously thought (2, 3, 13). The *B. oleracea* genome, along with several close relatives, including *Brassica rapa* (turnip) and *Brassica napus* (canola), has been the subject of intense study because of the agricultural importance of the group. The whole-genome triplication event that likely facilitated the origin of these crops was dated previously at 14 Mya (32). Our age estimate for the triplication is centered at 22.5 Mya (95% HPD: 28.3–15.6; *Table 1*).

Our data suggest that the Core Brassicales stem group (SG) originated around 71.3 Mya (95% HPD: 83.2–59.7), shortly before the end of the Cretaceous, when SG Capparaceae split from SG Brassicaceae and SG Cleomaceae (Fig. 2 and *Table 1*). Furthermore, using LASER (33), we detected a twofold shift in net diversification rate along the branch to core Brassicales when extinction rates are low or a 10-fold shift when extinction rates are high (*Table S5*). SG Brassicaceae split from SG Cleomaceae at the Cretaceous–Paleogene boundary (KPB; ≈ 65.5 Mya) (*Table 1* and Fig. 2). Thus, stem members of all three families may have survived the KPB mass extinction event before their subsequent radiations in the Eocene (Fig. 2 and *Fig. S4*). A pattern of origin before the KPB and radiation afterward has been noted in animal (34) and other plant lineages (35), although the lag time before recovery is greater in Brassicaceae than has been inferred in the other groups (35).

A great deal of interest has centered on the timing of WGD events in the history of *A. thaliana*, particularly on the most recent of these events, the α WGD, estimated to have occurred between 100 and 20 Mya (11, 36–39), and the β WGD, estimated to have occurred between 235 and 112 Mya (36, 38, 39) (Fig. 2). Recent lines of evidence place the α WGD within Brassicales (12), perhaps within Brassicaceae (26, 27, 36, 40), and place the β WGD within Brassicales (12, 40). These placements have fueled speculations that genome-doubling events are linked with diversification in Brassicaceae (13, 41) and with survival of the KPB mass extinction event (24). These hypotheses are attractive, because WGD may result in the colonization of new habitats (42), major ecological transitions (43), invasiveness (44), and the generation of morphological novelty (45). Our chronogram (Fig. 2 and *Fig. S4*) provides the framework for testing these hypotheses. For example, if the β WGD occurred around 70 Mya, placing it on the stem lineage of Core Brassicales (Fig. 2), we would expect to find that paralogs stemming from this event are shared by members of Brassicaceae, Cleomaceae, and Capparaceae but not by other families in Brassicales. If a WGD maps to this branch, the locus

Table 1. Comparison of published dates with those estimated in this study using Bayesian inference in BEAST on the *ndhF* + *PHYA* Brassicales chronogram

Node age	Koch et al. 2000 (2), 2001 (3) synonymous substitution rate, single fossil constraint (Mya)	Franzke et al. 2009 (27) BEAST, <i>nad4</i> , single secondary constraint (Mya)	Couvreur et al. 2010 (13) BEAST, supermatrix, single fossil constraint (Mya)	This study BEAST, <i>ndhF</i> + <i>PHYA</i> , four fossil constraints (Mya)
<i>Arabidopsis</i> CG	5.8–3	ND	ND	17.9 – 13.0 – 8.0
Brassicaceae CG (whole genome triplication)	14	ND	ND	28.3 – 22.5 – 15.6
Lineage I CG	19–13	19.0 – 8.0 – 0.5	36.1 – 27.3 – 18.2	42.8 – 35.6 – 28.5
Lineage II CG	18.1	ND	37.2 – 28.2 – 18.1	37.8 – 30.8 – 23.7
Lineage III CG	ND	ND	29.3 – 21.4 – 14.8	42.7 – 35.5 – 28.3
<i>Arabidopsis</i> – <i>Brassica</i> MRCA	21–16	ND	≈ 34*	50.7 – 43.2 – 36.6
Core Brassicaceae CG	ND	28.0 – 11.0 – 1.0	42.8 – 32.3 – 20.9	54.3 – 46.9 – 39.4
Brassicaceae CG	60–30	35.0 – 15.0 – 1.0	49.4 – 37.6 – 24.2	64.2 – 54.3 – 45.2
Brassicaceae–Cleomaceae MRCA	ND	35.0 – 19.0 – 1.0	ND	76.5 – 64.5 – 54.4
Core Brassicales CG	ND	ND	ND	83.2 – 71.3 – 59.7

Ages in bold are optimal reconstructions and are bracketed by upper and lower bounds of the 95% HPD interval. CG, crown group; MRCA, most recent common ancestor; ND, not determined.

*Approximate mean date inferred from chronogram and for which HPD data were not available (13).

of the shift in net diversification rate is consistent with a link between genome doubling and survival across the KPB mass extinction and perhaps to subsequent diversification.

Despite the interest in genome doubling and its effects, it is important to consider other factors that might promote diversification. Ehrlich and Raven (46) cited the interaction between Brassicales and pierid butterflies as a prime example of coevolution between plants and insects. Pierids specializing on Brassicales produce nitrile-specifier protein essential for detoxification of glucosinolates, a chemical defense found almost exclusively in Brassicales (10). This key innovation facilitated a host switch onto Brassicales, dated at ≈85 Mya (10, 47), and was followed by the subsequent diversification of pierids (10). The locus of the shift in Brassicales net diversification rate and the timing of the pierid butterfly radiation suggest that they codiversified (Fig. 2 and Fig. S4), supporting a central tenet of coevolutionary theory (46). However, the extent to which diversification in Brassicales may have driven pierid diversification, and vice-versa, requires further study.

Carefully designed comparative studies reveal the processes that generate and maintain biodiversity. In this study we identify a major shift in diversification rate that may correlate with survival across and diversification following the KPB mass extinction event, WGD events, and the evolution of pierid butterflies. Our results suggest a number of ways in which our understanding of the history of CG *Arabidopsis* should be revised, notably that it evolved during a period of warming rather than cooling and that genome structure and developmental processes have been slower to evolve than has been appreciated. The evolutionary history of *A. thaliana* and its neighborhood described here will allow more precise application of our understanding of this model organism to other flowering plants, land plants, and all eukaryotes.

Materials and Methods

GenBank accession numbers, specifics of the evaluation of published fossils, and detailed methods can be found in *SI Materials and Methods*.

Phylogenetic Inference. We used RAxML (48) to infer phylogeny in Brassicales from three alignments. The plastid *ndhF* alignment comprised 2,067 nucleotide sites representing 170 species; the nuclear *PHYA* alignment comprised 1,824 nucleotide sites representing 139 species; the combined alignment comprised 3,798 nucleotide sites from 177 species (Table S2). Analyses assumed a general time-reversible model of sequence evolution, with γ -distributed rate heterogeneity; partitions in the combined matrix were allowed to evolve independently. The topology and clade support in

the resulting trees are consistent with previously published phylogenetic estimates using *ndhF* (17) and other plastid (49) and nuclear (3, 22, 50, 51) markers.

Ultrametric Tree and Divergence Date Estimation. We calculated divergence dates for Brassicales from all three alignments using r8s v. 1.71 (30) and BEAST v. 1.5.3 (28). These methods account in different ways for variation in substitution rates among branches on the tree. We used r8s to explore the impact of placing the *Dressiantha* constraint at different nodes along the Brassicales backbone; potential placements were inferred in analyses of combined *ndhF* and morphological data (details are given in *SI Materials and Methods*). Because the placement of *Dressiantha* had little impact on age estimates within Core Brassicales (Table S3), we used a single placement in our BEAST analyses.

We allowed BEAST to infer topology, branch lengths, and dates for *ndhF* and combined data. For the *PHYA* data, we fixed the tree to that used in the r8s analyses and allowed BEAST to alter branch lengths while inferring dates (*SI Materials and Methods*). We used a uniform distribution for all three fossil calibrations with the lower hard bound of the distribution set to the youngest age of the fossil (see text, *SI Materials and Methods*, and Table S1) and the upper hard bound set to the first fossil record for eudicot pollen (125 Mya) (52) for *D. bicarpellata* and *Akania* sp. For other fossil calibrations the upper hard bound was determined by the age of other fossils used in the analysis (*SI Materials and Methods*). BEAST runs of 3×10^7 generations, saving data every 1,000 generations, produced 30,000 estimates of dates under a Yule speciation prior and an uncorrelated relaxed clock (28) for the single-gene datasets. Convergence statistics for each single-gene run were analyzed in Tracer, resulting in 27,000 post-burn-in trees. BEAST runs of 6×10^7 generations, saving data every 1,000 generations, produced 60,000 estimates under a Yule speciation (53) prior and uncorrelated relaxed clock (28) for the combined data. Also, for these data, we resampled at a lower frequency using LogCombiner v. 1.5.3 (28), resulting in a tree file with 30,000 trees. We used TreeAnnotator v. 1.5.3 (28) to produce maximum clade credibility trees from the post-burn-in trees and to determine the 95% probability density of ages for all nodes in the tree (Figs. S2–S4).

Shifts in Net Diversification Rate. We used LASER v. 2.2 (33) to test for significant shifts in net diversification rates over the history of Brassicales. Clade diversity for each group was determined from the Angiosperm Phylogeny Website (54). We calculated the likelihoods of our phylogenetic data under a model of constant diversification rate over time and under a two-rate model in which the diversification rate is permitted to change (Table S5). Because analyses of diversification rate can be sensitive to extinction, we determined whether our results were sensitive to differences in extinction fraction by reanalyzing the data under low ($a = 0$) and high ($a = 0.99$) extinction (Table S5).

ACKNOWLEDGMENTS. We thank Elisabeth Wheeler for her expertise in evaluating *Cappariodoxylon holleisii* using Inside Wood (<http://insidewood.lib.ncsu.edu>), Aleksej Hvalj (Komarov Botanical Institute, Russian Academy of Sciences) for help identifying paleobotanical references from Russian

scientific literature, Jocelyn Hall (University of Alberta, Edmonton, AB, Canada) for DNA material of Brassicales species, and Paul Forster (Queensland Herbarium, Brisbane, QLD, Australia) and David Woodlee (Burringbar

Rainforest Nursery, Burringbar, NSW, Australia) for plant material of *Akania bidwillii*. This work was supported by a Mercer Postdoctoral Fellowship (to M.A.B.) awarded by the Arnold Arboretum of Harvard University.

- Berr A, et al. (2006) Chromosome arrangement and nuclear architecture but not centromeric sequences are conserved between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Plant J* 48:771–783.
- Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabidopsis*, and related genera (Brassicaceae). *Mol Biol Evol* 17:1483–1498.
- Koch M, Haubold B, Mitchell-Olds T (2001) Molecular systematics of the Brassicaceae: Evidence from coding plastidic *matK* and nuclear *Chs* sequences. *Am J Bot* 88:534–544.
- Sherman-Broyles S, et al. (2007) *S* locus genes and the evolution of self-fertility in *Arabidopsis thaliana*. *Plant Cell* 19:94–106.
- Foxe JP, et al. (2009) Recent speciation associated with the evolution of selfing in *Capsella*. *Proc Natl Acad Sci USA* 106:5241–5245.
- Ha M, Kim ED, Chen ZJ (2009) Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc Natl Acad Sci USA* 106:2295–2300.
- Seoighe C, Gehring C (2004) Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends Genet* 20:461–464.
- Ziolkowski PA, Kaczmarek M, Babula D, Sadowski J (2006) Genome evolution in *Arabidopsis/Brassica*: Conservation and divergence of ancient rearranged segments and their breakpoints. *Plant J* 47:63–74.
- Felippes FF, Schneeberger K, Dezulian T, Huson DH, Weigel D (2008) Evolution of *Arabidopsis thaliana* microRNAs from random sequences. *RNA* 14:2455–2459.
- Wheat CW, et al. (2007) The genetic basis of a plant-insect coevolutionary key innovation. *Proc Natl Acad Sci USA* 104:20427–20431.
- Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res* 13:137–144.
- Ming R, et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452:991–996.
- Couvreur TLP, et al. (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol Biol Evol* 27:55–71.
- Becker HF (1961) *Oligocene Plants from the Upper Ruby River Basin, Southwestern Montana*. Geological Society of America Memoir (Geological Society of America), Vol 82, pp 1–127.
- Wing SL (1987) Eocene and Oligocene floras and vegetation of the Rocky Mountains. *Ann Mo Bot Gard* 74:748–784.
- Manchester S, O'Leary E (2010) Phylogenetic distribution and identification of fin-winged fruits. *Bot Rev* 76:1–82.
- Beilstein MA, Al-Shehbaz IA, Kellogg EA (2006) Brassicaceae phylogeny and trichome evolution. *Am J Bot* 93:607–619.
- Appel O, Al-Shehbaz IA (2002) Cruciferae. *Families and Genera of Vascular Plants*, ed Kubitzki K (Springer, Berlin), Vol 5, pp 75–174.
- Selmeier A (2005) *Capparidoxylon holleisii* nov. spec., a silicified *Capparid* (Capparaceae) wood with insect coprolites from the Neogene of southern Germany. *Zitteliana* 45:199–209.
- Gandolfo MA, Nixon KC, Crepet WL (1998) A new fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (Capparales). *Am J Bot* 85: 964–974.
- De Craene LPR, Haston E (2006) The systematic relationships of glucosinolate-producing plants and related families: A cladistic investigation based on morphological and molecular characters. *Bot J Linn Soc* 151:453–494.
- Beilstein MA, Al-Shehbaz I, Mathews S, Kellogg E (2008) Brassicaceae phylogeny inferred from phytochrome *A* and *ndhF* sequence data: Tribes and trichomes revisited. *Am J Bot* 95:1307–1327.
- Ermolaeva MD, Wu M, Eisen JA, Salzberg SL (2003) The age of the *Arabidopsis thaliana* genome duplication. *Plant Mol Biol* 51:859–866.
- Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc Natl Acad Sci USA* 106:5737–5742.
- Henry Y, Bedhomme M, Blanc G (2006) History, protohistory and prehistory of the *Arabidopsis thaliana* chromosome complement. *Trends Plant Sci* 11:267–273.
- Schranz ME, Mitchell-Olds T (2006) Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18:1152–1165.
- Franzke A, German D, Al-Shehbaz IA, Mummenhoff K (2009) *Arabidopsis* family ties: Molecular phylogeny and age estimates in Brassicaceae. *Taxon* 58:425–437.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
- Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Mol Biol Evol* 19:101–109.
- Sanderson MJ (2003) r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.
- Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292:686–693.
- Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassicaceae. *Genome Res* 15:516–525.
- Rabosky DL (2006) LASER: A maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evol Bioinf Online* 257–260.
- Montgelard C, Forty E, Arnal V, Matthee CA (2008) Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. *BMC Evol Biol* 8:321.
- McElwain JC, Punyasena SW (2007) Mass extinction events and the plant fossil record. *Trends Ecol Evol* 22:548–557.
- Bowers JE, Chapman BA, Rong JK, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438.
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 99: 13627–13632.
- Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. *Science* 290:2114–2117.
- Barker MS, Vogel H, Schranz ME (2009) Paleopolyploidy in the Brassicales: Analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biol Evol* 1:391–399.
- Soltis DE, et al. (2009) Polyploidy and angiosperm diversification. *Am J Bot* 96: 336–348.
- Rieseberg LH, et al. (2007) Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129:149–165.
- Rieseberg LH, et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211–1216.
- Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proc Natl Acad Sci USA* 97:7043–7050.
- Levin DA (1983) Polyploidy and novelty in flowering plants. *Am Nat* 122:1–25.
- Ehrlich P, Raven P (1964) Butterflies and plants: A study in coevolution. *Evolution* 18: 586–608.
- Braby MF, Vila R, Pierce NE (2006) Molecular phylogeny and systematics of the Pieridae (Lepidoptera: Papilionoidea): Higher classification and biogeography. *Zool J Linn Soc* 147:238–275.
- Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Koch MA, et al. (2007) Supernetwork identifies multiple events of plastid *trnF*(GAA) pseudogene evolution in the Brassicaceae. *Mol Biol Evol* 24:63–73.
- Bailey CD, et al. (2006) Toward a global phylogeny of the Brassicaceae. *Mol Biol Evol* 23:2142–2160.
- Galloway GL, Malmberg RL, Price RA (1998) Phylogenetic utility of the nuclear gene arginine decarboxylase: An example from Brassicaceae. *Mol Biol Evol* 15:1312–1320.
- Brenner GJ (1996) Evidence of the earliest stage of angiosperm pollen evolution: A paleo-equatorial section from Israel. *Flowering Plant Origin, Evolution and Phylogeny*, eds Taylor DW, Hickey LJ (Chapman and Hall, New York), pp 91–115.
- Yule GU (1924) A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis. *Philos Trans R Soc Lond B Biol Sci* 213:21–87.
- The Angiosperm Phylogeny Group (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot J Linn Soc* 141:399–436.